On the Regulation of Social Norms*

Dorothea Kübler†

Mai 5, 2000

Abstract

A model is developed to understand how norms can be influenced by "norm entrepreneurs", e.g. lawmakers, government agencies, unions etc. Two instruments of influencing the dynamics of norm-following behavior are analyzed, namely transforming the (monetary) incentives and changing the meaning or the reputational value of following a norm. First, incentives can be introduced (e.g. fines or subsidies imposed by government agencies) to violate existing norms or follow a new code of behavior. Second, actors can be convinced by norm entrepreneurs, e.g. using moral suasion, that following the existing norm is inappropriate or that following a certain new norm is appropriate. Both forms of norm regulation are incorporated into Akerlof's model of social custom (1980) in order to derive the comparative static properties of norm destruction and norm creation for different types of norms.

Keywords: social norms, regulation, conformity

*Thanks to Iris Bohnet, Steffen Huck, Wieland Müller, Jörgen Weibull, Georg Weizsäcker, Bengt-Arne Wickström, and participants of the workshop on "Economics and Social Norms" at Stockholm University. Financial support from Deutscher Akademischer Austauschdienst (DAAD) and Deutsche Forschungsgemeinschaft (SFB 373) is gratefully acknowledged. Part of the paper was written while I was visiting U.C. Berkeley in 1997/98 and KSG, Harvard University in 1999/2000. My thanks also go to the University of Bergen for its hospitality and to NFR-Ruhrgas for funding.

†Institut für Wirtschaftstheorie I, Humboldt-Universität Berlin, Spandauer Str. 1, D-10178 Berlin, Germany, kuebler@wiwi.hu-berlin.de.

We change our customs more quickly than our forefathers did, and we are more conscious of our changes and more willing to convert our customs into legal enactments, and to make them uniform. (p.638)

(...) in the absence of scientific history, short–lived man has little better means of ascertaining whether custom is quietly changing, than the fly, born today and dead tomorrow, has of watching the growth of the plant on which it rests. (p.640)

(Alfred Marshall, Principles of Economics, 8th ed., London: Macmillan, 1961)

1 Introduction

Social norms have recently gained much attention from economists as an important driving force of individual behavior. Legal scholars are primarily interested in norms in relation to formal rules such as laws, rights, or statutes.[1] Among other questions the legal discussion focuses on how formal rules influence norm–guided behavior indirectly by changing the social norms themselves and on whether norms can and should serve as a source of or a substitute for legal rules.[2] Also, norms are often considered as an efficient means to guide behavior because they do not require enforcement agencies, jails etc., but are self-enforcing.[3] For a long time, economists have treated norms as exogenous or, as Marshall puts it, as so sticky that changes are barely observable.[4] But norms and customs undergo changes, and at an increasing pace as Marshall already noted at the end of the 19th century. More recently, the processes underlying norm formation have been formally investigated by economists, e.g. by analyzing the evolution of norms and

[1] See for example the papers presented at the Symposium on Law, Economics, and Norms at the University of Pennsylvania Law School, published in the University of Pennsylvania Law Review, 1996, vol. 144.

[2] See Cooter and Ulen (1997), p.377–78, who mention ethical standards developed by formal or informal networks of professionals as sources for formal law.

[3] For example, the efficiency of norms in regulating crime is stressed by those who advocate shaming punishments. In California, a person who was caught shoplifting was convicted to wear a T–shirt declaring "I am a Thief". See *The New Yorker*, Oct 20 & 27, 1997, p.174.

[4] For a discussion and critique of this issue see Basu, Jones, and Schlicht (1987). Schlicht (1998) provides an extensive discussion of changes of custom resulting from fuzziness of customs, self-interested behavior, and changes in the incentive structure.

standards in certain environments.[5] In contrast to this literature, this paper tries to model deliberate attempts by certain members or groups of society to change the existing set of norms. Concluding his paper on the evolution of social norms, Peyton Young (1998) writes:

> (...) social change occurs for a complex variety of reasons. It is driven to some degree by the opinions and actions of influential people (role models) who precipitate change because they are widely noticed and imitated. Furthermore, changes in one sphere of interaction may have important spillover effects on other spheres and are related to broad underlying trends in society. (...) we have abstracted away from such complications.

In the following it is studied how a new code of behavior can be established or an old existing code can be destroyed by actions of "norm entrepreneurs" such as unions, the church, government agencies, NGO's and interest groups or single influential individuals.[6] The model below is an attempt to formalize a number of issues of the informal legal discussion of norms and to understand the dynamics of social customs in the presence of a norm–regulating agency.

Akerlof (1980) seeks to explain why social norms involving pecuniary disadvantages for those adhering to them do not necessarily disappear. He shows that norms can, but do not necessarily erode if there are a number

[5] For an example of the evolutionary approach applied to norms see Young (1998) who shows that evolutionary forces may lead to norms implying efficient and more or less egalitarian outcomes. Huck (1998) studies the impact of legal rules on the evolution of preferences.

[6] There is a difference between government interventions to change the set of norms and actions of private parties. For the government, there is a "liberal constraint on social meaning making" (Lessig, 1995) as people have strong feelings against manipulations of this kind by the state, reminding them of totalitarian propaganda so pervasive in the past century. Therefore, the government is often more successful in conveying certain messages when it does so indirectly, e.g. by letting school teachers speak out against racial discrimination or doctors in favor of using condoms to prevent the spreading of AIDS. But a certain asymmetry in the evaluation of government actions can be observed: If the status quo or reigning orthodoxy is supported, there is not much resistance, but any attempt to challenge or even change the dominant view is seen as problematic. Also, corporate propaganda is ubiquitous (see Corneo and Jeanne (1997) who model a monopolist creating a "consumption norm" for his product). Although there is and should be a difference between the appropriateness of government and corporate strategies to influence social norms, it does not seem obvious why the government should have no role in the process at all.

of people who disobey the code of behavior, which in turn undermines the beliefs of those still following the norm. Thus, disobedience and erosion of the belief in a norm may reinforce each other until the social code dies out. However, this process need not take place even if the norm is disadvantageous for everybody in the group because not following the norm may involve a loss of reputation. It is, among other things, the wish to conform to existing customs that drives individual behavior.

Two aspects of Akerlof's model are worth emphasizing. First, a norm is not defined as a certain behavior followed by a number of people as in the literature on evolutionary game theory. Rather, it is a moral expectation shared by a group of people, entailing social stigmatization or at least moral indignation aimed at those who deviate. Second, in Akerlof's model the survival of a norm does not depend on the material success or some other inherent property of following the norm, but rather on the number of followers at the initial stage.[7] For social norms, this seems to be a more adequate approach than looking at the actual payoff received from following a certain code of behavior because it is by no means clear that norms always promote either self–interest, common interest, or genetic fitness. And even if the benefits of a norm are collective in nature, individual reinforcement does not necessarily work in the right direction.[8]

Norms are constituted by common understandings and expectations of a group of individuals. In order to change these taken-for-granted expectations and to establish a new code of behavior, it is typically not sufficient that one individual changes his or her behavior. Rather, a group of individuals must

[7]Several mechanisms can bring inefficient norms into existence: Technological, economic or social change might render previously useful norms inefficient, efficient norms do not develop immediately, but often with considerable time lags involved (see Posner (1996) who discusses Demsetz' famous article about fur trading native tribes in Canada and argues that the common pool problem due to trading furs arose much earlier than property rights developed, which led to a depletion of fur areas during that period), and there is no mechanism securing that an efficient norm develops whenever it is beneficial for all participants. Also, norms may develop on the basis of certain principles such as salience, prominence or analogy which do not guarantee, but are even unrelated to their optimality (see Sugden, 1989). And finally, if third parties who are affected do not participate in the norm–generating process, the resulting norm may be inefficient. For example, cartel rules or codes of business ethics usually do not reflect the interests of consumers (see Schäfer and Ott, 1993) .

[8]For an extended discussion along these lines see Elster (1989) who argues for the "autonomy of norms" (p.114).

embrace and follow the new rule to give it the status of a social norm. But to contribute to the public good of establishing a better norm (for example) is costly for individuals due to the old norm itself that stigmatizes any deviation from it. To put it more generally, a norm can solve a collective action problem, but for a norm to become effective (or to disappear), such a collective action problem has to be solved first. The norm has to be shared by a certain number of individuals to guarantee that its content is binding as it is enforced via social stigmatization or other informal sanctions. Solving this collective action problem is the task of norm regulation.

In Akerlof's world of persisting inefficiencies of norms, there is room for welfare improvement by norm regulation. In addition, interest groups may try to change a code of behavior that is disadvantageous to its members, even if the code is beneficial from a social welfare perspective. The question is how the problem of collective action can be solved by interventions of a central agency, e.g. lawmakers, non-government and government agencies, interest groups etc. Often norms are first attacked by interest groups or single persons acting as norm entrepreneurs. In some cases government agencies step in to administer the desired change that is blocked by problems of collective action or by opposing interest groups.[9] Also, unions have acted as norm entrepreneurs.[10]

In the model below, two possible instruments of influencing the dynamics of norm formation and erosion are analyzed.[11]

[9] Sometimes this intervention is not even necessary, because a norm is fragile enough to vanish after it is publicly criticized. However, it is an important step taken by the norm-entrepreneur to make this criticism common knowledge as people who want to break the code might only be willing to do so if they expect others to do so as well.

[10] To press for the 5–day work week in Germany, unions started an advertising campaign with a poster of a child saying: "On saturdays, daddy is mine!" (The German slogan was "Samstags gehört Papi mir!") in order to associate the 5-day work week with family friendly working conditions. Lessig (1995) describes this technique of social regulation as "tying".

[11] For a non–formal discussion of different techniques of norm regulation see Lessig (1995) and E. Posner (1996). In addition to the two mechanisms discussed below, which Posner calls norms–violation and norm–transformation, he also describes norm–circumvention as bargaining around inefficient norms by private parties. The government can facilitate norm–circumvention by lowering bargaining costs, e.g. refining property rights. Lessig distinguishes between behavioral and semiotic techniques of norm change which are then differentiated into the semiotic techniques of "tying" and "ambiguation" as well as the behavioral techniques of "inhibiting" and "inducing" a certain behavior.

First, incentives to follow or break a norm can be established by subsidizing, punishing, or prohibiting a certain behavior, for example norm erosion can be achieved by granting monetary rewards to those who deviate from a prevailing norm.[12] Changing the incentives to break a norm, i.e. a change in relative prices of norm following vs. norm breaking, is referred to as "regulation of motives". For example, a norm of hiring women only if there is no (equally qualified) man applying for the job can be discouraged by rewarding those employers who hire women with a better chance to receive government contracts. Or ecological norms could be strengthened by refunds for returned glass bottles.[13] Another example are labor supply norms for men and women. The German tax system gives tax breaks to one-breadwinner households, but not to households with two incomes and thus favors an asymmetric division of labor. Although the tax system is formally gender neutral, it is argued that it reinforces preexisting gender roles and discourages women from participating in the labor market.

The second technique to change norm-following behavior involves changing the content of the norm or convincing actors that following the norm is inappropriate. Here, the norm entrepreneur tries to change the reputational value of the norm instead of giving incentives to break a widely accepted norm as with regulation of motives. This is modelled by assuming that the social pressure exercised by other norm followers becomes weaker when there is a norm entrepreneur who discredits the norm. Or in other words, deviation from the norm is vindicated by the norm entrepreneur's action, which is referred to as "regulation of meaning". For example, the government can try to convince actors of the inefficiency or inappropriateness of a norm by performing symbolic acts itself. If the government action has some moral weight of its own, the force of the norm, its reputational value, declines. Returning to the above example of the norm requiring gender–based hiring

[12] Hirschman (1985) argues that raising the cost of an action does not stigmatize it in the same way as making it unlawful. A law prohibiting an action without monetary sanctions would fall under the category of regulation of meaning (discussed in the next paragraph), whereas monetary sanctions change the incentive structure.

[13] However, simple monetary incentives can also be ineffective. Consider the case where norms against genetically manipulated food exist, e.g. it is considered risky to give this food to children, although using the technology is legal. Suppose that prices for genetically manipulated food are lower, either due to subsidies or because of a more efficient production. Nevertheless, the existence of the norm and the slow process of norm change may cause losses to firms using the technology. Finally, the norm may erode or the technology is driven out of the market.

decisions, state agencies can use non–discriminatory hiring practices or even affirmative action programs to express their criticism of the existing norm.[14] Or they may attempt to change norms via education programs, for example by adjusting textbooks or instructing teachers to question certain codes of behavior. The behavioral code of "safer sex" was established as a response to the spreading of AIDS and a dysfunctional system of norms governing sexual relations, mainly by extensive advertising campaigns associating condom use with responsible behavior. In the domain of free speech, civil disobedience and free expression can be rewarded for example by creating prizes with independent juries to honor the courage to stand up for one's beliefs and resist the norm "don't stick your neck out". These norm–transforming acts imply that even if the number of people following the norm is still rather high, belief in the norm deteriorates and, in the long run, less people are going to abide by it.

In the following, both techniques of norm regulation are incorporated into a model of social custom in order to analyze their short– and long–run effects on the proportion of norm followers. For a general reputational value function it can be shown that when most people follow a certain norm, changing the incentive structure has a bigger marginal effect on the number of norm followers than changing the reputational value of the norm.

Two types of norms characterized by different reputational value functions are analyzed. If the reputational value of a norm or the social pressure exercised by it increases sharply, once a critical number of norm followers is reached, this is called a bandwagon norm. Many norms seem to exhibit this property, for example the social meaning attached to smoking. As long as a large enough group of people smokes, smoking is considered entirely acceptable. But when the group of smokers is small in a society, nonsmokers might feel more justified in imposing restrictions on smoking. An inefficient bandwagon norm, adhered to by a large proportion of people, can best be weakened by introducing incentives to break it. On the other hand, to create a bandwagon norm, a group of sufficient size must simultaneously adopt the new norm to get the bandwagon started. For example to establish recycling as a norm, a critical mass of people must start returning bottles etc. in order to convince others to do the same.

Snob norms refer to norms that yield most reputation to its followers when only a small group of people follow it. It is shown that increasing the

[14] See Sunstein (1996) for a discussion of the expressive role of law.

incentives to break a snob norm increases its reputational value. For example when cocaine gets more expensive, cocaine consumption might become even more prestigious for those who can still afford it. Thus, to substantially decrease the number of cocaine users, the price of cocaine must be very high to compensate for these reputation gains. Policies that try to convince people of the dangers of cocaine consumption can be cheaper and more effective. Thus, in contrast to bandwagon norms, regulating the social meaning of snob norms is more effective to destroy such norms than changing the incentive structure.

The model below does not specify a particular game, but assumes that an inefficient norm exists such that absent social stigmatization, it is in everybody's interest to break the norm. And, by a change of signs, the model addresses the case where an efficient norm is absent. The plan of the paper is the following: In Section 2, the model both for the short and the long run is introduced and the equilibria are characterized. For a simple linear reputational value function, the equilibrium dynamics are derived and the comparative static properties of the two policies of norm regulation are analyzed. In Section 3, the assumption of a specific function for the reputational value of norm following is relaxed, leading to a general comparative static result and to the analysis of bandwagon and snob norms. Section 4 concludes the paper.

2 The Model

First, let me reformulate the model of Akerlof (1980) for the purpose of this paper. Consider a situation in which a norm is clearly defined and well known to everybody. Suppose for simplicity that it is only possible to either follow the norm or not. First, take the case of an inefficient norm that prevents people from consuming a private good yielding utility $y > 0$. The index variable A takes on the value of 0 if the individual follows and 1 if he or she breaks the code. Furthermore, μ represents the fraction of the population of size normalized to 1 that believes in the norm, which is fixed in the short run. The taste parameter a_i follows a uniform distribution on the unit interval $I = [0, 1]$. It expresses how much the individual cares about his or her reputation, i.e. about what others believe to be the right behavior. Finally, $m_g > -y$ represents the incentive to disobey the code introduced by the norm entrepreneur, e.g. a monetary reward for breaking the norm.

Individual utility is given by

$$U(a_i) := (y - a_i\mu + m_g)A. \tag{1}$$

Each individual is assumed to maximize this utility function, which constitutes the short-run equilibrium.

Notice that the norm contradicts the individual's private interest as $y > 0$ (and $m_g = 0$ for the moment), i.e. absent reputational concerns ($a_i = 0$), a person would always break the code. The higher the proportion of believers in the code and the higher a_i, the greater the disutility from not following the code. It is assumed that those with the highest taste for reputation, a_i, are also the ones who believe in the social custom. Thus, the strength of the societal belief in a norm, μ, is the aggregate of those individuals who believe in the norm. Note that making the utility dependent on the proportion of believers in the norm instead of on the proportion of norm followers allows for inertia in the model. When the behavior of some people changes, norm following behavior of others does not change immediately, but only after the societal belief in the norm has changed, too. Moreover, distinguishing between the number of followers and believers allows for a richer description of the reputational value of norm following, which will be introduced below. However, for simplicity it is assumed that believers and non–believers have the same utility function.[15]

The model also encompasses the case where nobody follows a norm because it is individually costly and there is no social pressure to follow it. In these instances, a new norm can be established, for example because it is collectively beneficial or a certain group of people is favored by it or future generations will profit from it. In this case, $m_g < 0$ and those who do not follow the code are punished, with the maximum punishment smaller than the private benefit from breaking the code.

Now consider the long run. If some people disobey a given norm, this undermines belief in the norm. The reputational value $v(\cdot)$ of a norm which depends on the proportion of norm followers, denoted by x, drives the dynamics of belief in the code.[16] In particular, if the reputational value

, of a code differs from the current proportion of believers, this proportion μ will increase or decrease at rate $\delta \in (0, 1)$ until they are equal in equilib-

[15] The model can be extended in a straightforward manner to allow for believers suffering more from breaking the code than non–believers, due to cognitive dissonance for example. See Naylor (1989).

[16] See Corneo and Jeanne (1997) for a discussion of the reputational value of a norm.

rium. And remember that $v(\cdot)$ adjusts to a change in μ as the number of people following the norm, x, depends on μ as specified in the utility function (1). The long–run dynamics of the model are described by the following equation:

$$\frac{\partial \mu}{\partial t} = \delta(v(x_t, x_g) - \mu_t). \tag{2}$$

The function characterizing the long–term reputational value of a norm, $v(x_t, x_g)$, defined on $[0, 1]$, depends not only on the number of individuals following it in period t, x_t, but also on the norm entrepreneur's action, x_g, directed at transforming the meaning of the norm. Assume that $\partial v / \partial x_g \geq 0$.[17] Regulation of meaning influences individual actions only in the long run, changing the set of long–run equilibria. Intuitively, if x_g is reduced for a given number of norm followers x_t, fewer people believe in the norm, i.e. μ_t is lower than without the government policy.
To obtain closed form solutions, in the next sections of the paper the reputational value is assumed to take the functional form

$$v(x_t, x_g) = \left\{ \begin{array}{cc} 0 & \text{if } x_t + x_g < 0, \\ x_t + x_g & \text{if } 0 \leq x_t + x_g \leq 1, \\ 1 & \text{otherwise.} \end{array} \right\}$$

In Section 3 this assumption is relaxed.
The relationship between short and long run is the following: In the beginning of each period t, the proportion of believers, μ_t, is fixed and individuals decide whether to follow the norm or not, given this proportion of believers, by maximizing their respective utility function (1). Between periods t and $t + 1$, beliefs of the new period are formed according to Equation (2).

2.1 Short-run equilibrium

An individual i's optimal decision is

$$A = \left\{ \begin{array}{ll} 0 \text{ if } & a_i > \frac{y + m_g}{\mu} \\ 1 & \text{otherwise.} \end{array} \right\}$$

[17] The effort to transform the norm may also be resisted, e.g. because people believe that a certain issue should not be regulated by the government or because the norm entrepreneur does not use convincing arguments in his campaign. In this case, the policy has either no effect or is even counterproductive, $\frac{\partial v}{\partial x_g} \leq 0$.

Suppose there exists a critical parameter a_c at which $y - a_c\mu + m_g = 0$. Then, the proportion of norm followers at time t is given by

$$x_t = \left\{ \begin{array}{ll} 1 - a_c = \frac{\mu_t - (y+m_g)}{\mu_t} & \text{if } \mu_t > y + m_g \\ 0 & \text{otherwise.} \end{array} \right\} \tag{3}$$

The number of norm followers is a piecewise continuous decreasing function in the private benefit from breaking the norm, y, and the policy variable m_g. By introducing incentives to break the norm, m_g is increased from 0 to $m_g > 0$. If these incentives to break a norm change, the loss of reputation due to violating an existing norm, $a_i\mu$, remains unchanged, but every individual gets some monetary reward for breaking the code. Some types a_i might change their behavior in response. Thus, the strength of the individual adherence to a particular code of behavior depends on how the agency chooses its actions with respect to that code.[18]

2.2 Long-run equilibrium

In the long run, not only the number of followers of the norm is endogenous, but also the strength of the societal belief in the norm. The evolution of μ_t is specified by Equation (2). For stationarity, i.e. $\partial\mu/\partial t = 0$, it is necessary and sufficient that $\mu_t = x_t + x_g$. Moreover, the number of norm followers x_t is directly related to the number of believers μ_t by Equation (3).

Proposition 1 *1. If the monetary incentives to break the norm are weak, i.e. $y + m_g \leq x_g$, the only stable equilibrium is $x^* = 1$.*

2. If the monetary incentives to break the norm are of intermediate strength, i.e. $\frac{1}{4}(1 + x_g)^2 > (y + m_g) > x_g$, two stable long–run equilibria exist:

$$x^* = 0 \qquad \textit{and} \tag{4}$$

$$x^* = \frac{1}{2}(1 - x_g + [(1 + x_g)^2 - 4(y + m_g)]^{1/2}). \tag{5}$$

[18] Analogously, individual i decides to follow a new norm although it is costly to him or her if

$$y - a_i\mu + m_g < 0.$$

Thus, by decreasing m_g sufficiently, the norm entrepreneur can induce more types to follow the new code.

3. *If the monetary incentives to break the norm are strong, i.e.* $(y+m_g) \geq \frac{1}{4}(1+x_g)^2$, *the only stable equilibrium is* $x^* = 0$.

Proof Solving Equation (3) together with $\mu_t = x_t + x_g$ for x_t yields

$$x^*_{1,2}(m_g, x_g) = \frac{1-x_g}{2} \pm \frac{1}{2}[(1+x_g)^2 - 4(y+m_g)]^{1/2}. \qquad \blacksquare$$

This solution describes two equilibrium points, the upper one being stable, the lower unstable. If the short-run equilibrium schedule and the locus of long-run equilibrium points intersect twice, two long-run equilibria exist. If the convex schedule is just tangent to the locus of long-run equilibria, i.e. the expression in brackets is zero, only one equilibrium exists. If the short-run equilibrium schedule lies everywhere above the long-run schedule, no solution in real numbers exists and the only long-run equilibrium is $\mu_t = x_t + x_g = 0$. Finally, if the short-run equilibrium schedule lies everywhere below the long-run schedule, the only equilibrium is $x^* = 1$.

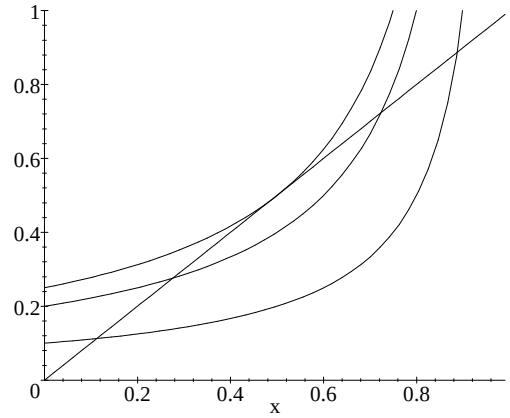

Figure 1: Equilibria and regulation of motives

The solution can be illustrated in an (x, μ)–diagram shown in Figure 1. Reformulating Equation (3) yields the short-run equilibrium curve described by

$$\mu_t = \left\{ \begin{matrix} \frac{y+m_g}{1-x_t} & \text{if} \quad 1 - x_t > y + m_g \\ 1 & \text{otherwise,} \end{matrix} \right\} \tag{6}$$

i.e. the number of believers is a piecewise continuous increasing function of the number of norm followers.[19] If $x_g = 0$, long-run stability is represented by the 45°–line. Together with the assumption that high a-types are believers in the norm, this means that when $x_g = 0$, all followers of a norm are also believers and vice versa. In the case of two long-run equilibria, the upper equilibrium is asymptotically stable whereas the lower equilibrium is unstable. This can be seen by investigating points on the short-run equilibrium curve above and below the $\mu_t = x_t + x_g$–schedule. If $\mu_t < x_t + x_g$, both μ_t and x_t fall according to equation (2). But below the schedule where $\mu_t > x_t + x_g$, both variables rise. Thus, if the number of norm followers is at or below the lower equilibrium, $x^* \leq \frac{1-x_g}{2} - \frac{1}{2}[(1+x_g)^2 - 4(y+m_g)]^{1/2}$, the norm erodes, while starting at any proportion of followers above it, the number of norm followers eventually reaches the stable long-run equilibrium. Note that $x^* = 1$ is a stable equilibrium if $y + m_g = 0$ and $x_g \geq 0$.

Destruction of an existing norm can be partial by pushing down the stable equilibrium and thus reducing the number of norm followers. Completely destroying an existing norm requires reaching the tipping equilibrium and thus necessitates a large enough change in policies. Similarly, establishing a new norm requires pushing the lower equilibrium down to $x_t = 0$ such that just one person following the code can set off the process of norm-generation, characterized by a growing number of believers and followers over time. Two different types of policy can bring about these changes.

2.3 Regulation of motives

Figure 1 represents the effect of regulating motives, i.e. giving individuals an incentive m_g to break or follow the norm.[20] Subsidizing norm-breaking behavior moves the short-run equilibrium schedule upwards and to the left and thus shifts the equilibria. In particular, the upper equilibrium moves downwards and the lower equilibrium upwards along the $\mu_t = x_t + x_g$–schedule. For example, as the number of norm followers in the upper equilibrium is de-

[19] Moreover, for $1 - x_t > y + m_g$, the short-run equilibrium curve is convex as

$$\frac{\partial \mu_t}{\partial x_t} = \frac{y + m_g}{(1 - x_t)^2} > 0 \quad \text{and} \quad \frac{\partial^2 \mu_t}{\partial {x_t}^2} = \frac{2(y + m_g)}{(1 - x_t)^3} > 0 \tag{7}$$

for $x_t \in [0, 1)$.

[20] For a similar treatment of the impact of incentives on social customs see Schlicht (1998), p. 46-49.

creased, this implies that in the long-run equilibrium the number of believers μ_t is reduced, too . If the two schedules do not intersect any longer, people do not want to follow the norm for any a_i and μ_t because the reward for breaking the norm becomes sufficiently large, and the norm vanishes. Therefore, taking the stable equilibrium with $x^* > 0$ as a starting point, the effect of creating incentives to break the code depends on the size of m_g. Either the proportion of followers and believers is simply moved down to the new stable equilibrium or the norm dies out completely.

Now consider a situation in which a norm is not followed by anybody, $x_t = 0$. Setting $m_g < 0$, i.e. introducing positive sanctions for norm followers, pushes down the tipping equilibrium. Once this unstable equilibrium reaches $x_t = 0$, a single norm follower is enough to induce a norm change leading to the only remaining stable equilibrium with a positive proportion of norm followers.

2.4 Regulation of meaning

Regulating the meaning of a norm only affects long-run adjustments and equilibrium by reducing the reputational value of a norm for a given proportion of norm followers. A certain norm-guided behavior can be discredited by advertisement campaigns or symbolic acts of influential people, thereby reducing the social pressure to follow the norm exercised by norm followers. Thus, in equilibrium a given number of followers is associated with a smaller proportion of believers than without the norm–transforming policy.[21] Under the assumption that $v(x_t, x_g) = x_t + x_g$, meaning regulation amounts to a parallel shift of the $\mu_t = x_t + x_g$–schedule.

[21] For a more general discussion of moral suasion and education as instruments of economic policy see Frey and Kirchgässner (1994), p. 403.

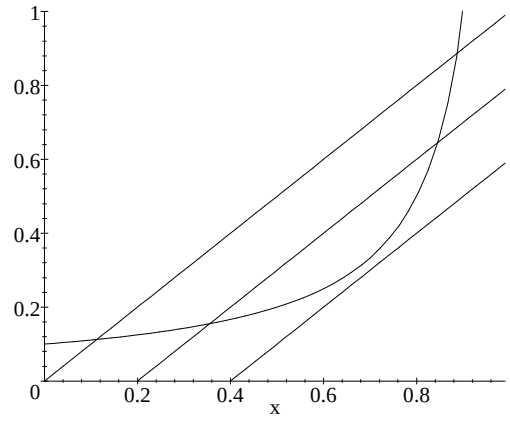

Figure 2: Regulation of meaning

Figure 2 depicts the effect of increasing or decreasing the reputational value of a norm for a given number of norm followers. Notice that the qualitative changes are the same as with regulating motives: After reducing $v(x_t, x_g)$ by setting $x_g < 0$, the upper equilibrium shifts downwards along the curved schedule while the lower unstable equilibrium shifts upwards. Again, depending on the size of x_g, norm–following behavior is either reduced or erodes completely. On the other hand, increasing x_g shifts the 45°–line upwards, thus decreasing the critical number of norm followers necessary to set off a process of norm formation. Simultaneously, the stable upper equilibrium moves further up.

2.5 Equilibrium dynamics

The equilibrium dynamics in terms of x_t are found by differentiating Equation (3) with respect to t and plugging in Equation (2):

$$\frac{\partial x_t}{\partial t} = \frac{\delta(x_t + x_g - \mu_t)(y + m_g)}{\mu_t^2} \quad \text{if} \quad \mu_t > 0. \tag{8}$$

(If $\mu_t = 0$, $x_t = 0$ is reached immediately by definition in Equation (3) as $y + m_g \geq 0$.) The larger y, m_g, x_g, and x_t, the faster x_t changes over time; that is, for example, the higher the utility from the private good, the faster the norm erodes. Similarly, if the number of norm followers increases as $x_t + x_g > \mu_t$, this increase is faster, the bigger y, m_g, x_g,and x_t.

Regulation of motives and meaning can both lead to a deterioration of the norm. However, the process leading to the long–run equilibria is different under each policy. Modifying the motivational structure immediately changes the behavior of those individuals for whom breaking the norm becomes attractive although all norm breakers incur a disutility of $-a_i\mu$ from a loss of reputation. In Figure 3, this is represented by the horizontal shift from the old equilibrium to a point on the new short-run equilibrium schedule with μ unchanged. This sets off the adjustment process leading to a new long-run equilibrium.[22]

Transforming the meaning or social value of a norm only affects individual behavior indirectly by reducing the proportion of believers and thus the reputational value of following the norm. The speed of belief in the norm adjusting to actual behavior and vice versa is parametrized by δ. If δ is small, regulation of meaning leads only to a slow change of norm–following behavior. But individuals do not experience any disutility from social sanctions for breaking a still powerful norm as in the case of regulation of motives where a small δ means high reputational losses for the norm breakers as the norm is sticky.

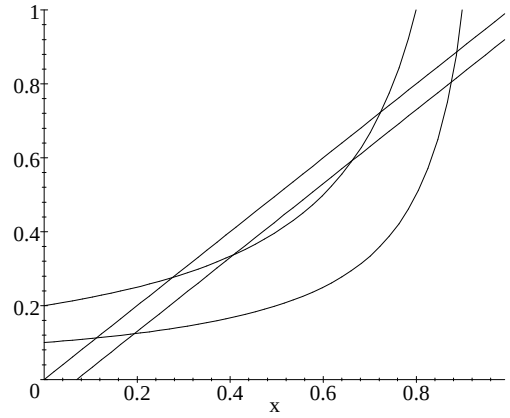

Figure 3: Cross effects, adjustment process

[22]For a related analysis of the direct and indirect effects of legal institutions see Huck (1998). In an evolutionary framwork he shows that punishments not only deter a certain action in the short run, but can also have long-run effects. For certain punishment levels, types who feel remorse are materially better off than others. Thus, the evolutionary process will lead to an increase in the proportion of types feeling remorse.

Figure 3 illustrates the adjustment process after increasing m_g and decreasing x_g simultaneously, starting in the upper stable equilibrium. By introducing incentives to break the norm, the number of norm followers jumps to the left immediately, represented by the long horizontal arrow. In the long run, this leads to a deterioration of beliefs in and obedience to the norm, moving along the short-run equilibrium curve according to the small arrows. The process ends at the stable point on the new long-run equilibrium line, which lies below and to the left of the stable point that would have been reached without a decrease in x_g.

2.6 Comparative statics

Investigating the marginal effects of both regulation techniques complements the graphical analysis performed in the subsection above. Consider a situation where a norm is followed by the same proportion of people in every period, i.e. (a) a stable long–run equilibrium with $x^* > 0$ or, (b) a stable equilibrium with no norm–following behavior, $x^* = 0$.

Proposition 2 *(a) Creating incentives to break an existing norm, i.e. raising m_g marginally, decreases the stable long–run equilibrium number of norm followers and believers x^* and μ^* by the same amount,*

$$\frac{\partial \mu^*}{\partial m_g} = \frac{\partial x^*}{\partial m_g} = -[(1 + x_g)^2 - 4(y + m_g)]^{-1/2} < 0. \tag{9}$$

(b) Creating incentives to follow a new norm by lowering m_g at $x^* = \mu^* = 0$ leads to a new equilibrium $x^* = 1$ iff

$$v(0, x_g) \geq \mu_t \Longleftrightarrow x_g \geq y + m_g. \tag{10}$$

Proof Straightforward. ■

As an increase in m_g results in a new stable equilibrium on the $\mu_t = x_t + x_g$–line, raising m_g has the same effect on the equilibrium number of followers and believers, x^* and μ^*. Intuitively, as the individual benefit from breaking the norm is increased, the former critical type a_c now wants to break the norm and the number of followers decreases, leading to an equal decrease in believers in the long run. In the stable equilibrium in which nobody follows the norm, a marginal change of m_g may have no effect as

the tipping point has to be reached to start the process of norm emergence. And in particular, if $x_g = 0$, m_g has to be lowered to $y = -m_g$ such that all types of individuals are just indifferent between following and not following the norm at $\mu_t = 0$. Similarly, destroying the norm completely necessitates reaching the tipping equilibrium, i.e. a change in policies is necessary such that $\frac{1}{4}(1 + x_g)^2 = (y + m_g)$.[23]

Now consider the effect of transforming the meaning of a norm, i.e. a variation of x_g:

Proposition 3 *(a) Discrediting a norm, i.e. reducing x_g, decreases the stable long–run equilibrium number of norm followers and believers x^* and μ^* in the following way:*

$$\frac{\partial x^*}{\partial x_g} = -\frac{1}{2} + \frac{1}{2}[(1 + x_g)^2 - 4(y + m_g)]^{-1/2} > 0 \tag{12}$$

$$\frac{\partial \mu^*}{\partial x_g} = \frac{1}{2} + \frac{1}{2}[(1 + x_g)^2 - 4(y + m_g)]^{-1/2} > 0. \tag{13}$$

(b) To establish a new norm, the raise in x_g must be sufficient to satisfy Inequality (10).

Proof Straightforward. ■

Summarizing these results, if the status quo is characterized by the stable long–run equilibrium with $x^* > 0$, increasing m_g has a stronger marginal effect than decreasing x_g on the number of norm followers.[24]

To evaluate a mix of both policies, the cross derivatives must be examined. They are given as

$$\frac{\partial^2 \mu^*}{\partial m_g \partial x_g} = \frac{\partial^2 x^*}{\partial m_g \partial x_g} = [(1 + x_g)^2 - 4(y + m_g)]^{-3/2}(1 + x_g) \tag{14}$$

[23] In the instable equilibrium, the marginal effect of m_g is

$$\frac{\partial \mu^*}{\partial m_g} = \frac{\partial x^*}{\partial m_g} = (1 - 4y)^{-1/2} > 0. \tag{11}$$

[24] This can be checked by solving the following inequality with $y \in (0, .25)$

$$-(1 - 4y)^{-1/2} < -\frac{1}{2} - \frac{1}{2}(1 - 4y)^{-1/2}.$$

.

for the upper stable equilibrium. As the term in brackets is positive when $x^* > 0$, the cross effect is positive. The marginal effect of m_g on μ^* and x^* is larger, the bigger x_g. If for example m_g is increased and x_g reduced to erode the power of the norm, the effect of motive regulation, m_g, is enhanced by the regulation of meaning, x_g. Thus, the two policies are complements, which is illustrated in Figure 3.

The question remains how general these comparative static properties are. This is investigated in the following section.

3 Generalized reputational value function

Some of the above results rely on the specific function for the reputational value of following a norm, $v(x_t, x_g) = x_t + x_g$. But many other specifications seem justifiable when thinking about examples in the real world. A simple modification is the case where the effect of regulating the meaning is proportional to the number of norm followers, $v(x_t, x_g) = x_t x_g$ with $x_g \geq 0$. Here, regulation of meaning has a greater effect, the more people follow the norm. But the reputational value can also be non-linear in x_t, for example if there is a critical number of norm followers necessary for others to feel forced to follow the norm as well. Which of the comparative static properties described above rely on the particular linear additive specification? The general results are summarized in

Proposition 4 *The marginal effects of m_g and x_g in the stable equilibrium with $x^* > 0$ are as follows:*

$$\begin{aligned} \frac{\partial x^*(m_g, x_g)}{\partial m_g} &= \frac{-\frac{1}{1-x^*}}{\frac{y+m_g}{(1-x^*)^2} - \frac{\partial v}{\partial x^*}} \\ \frac{\partial x^*(m_g, x_g)}{\partial x_g} &= \frac{\frac{\partial v}{\partial x_g}}{\frac{y+m_g}{(1-x^*)^2} - \frac{\partial v}{\partial x^*}}. \end{aligned}$$

Proof By implicit differentiation of $v(x_t(m_g, x_g)) \equiv \frac{y+m_g}{1-x_t(x_g, m_g)}$. ■

If the number of norm followers is high, i.e. x^* is close to 1, then regulation of motives is very effective at the margin. With the denominator of both

effects being equal, the comparison with the marginal effect of regulating the meaning of the norm depends critically on $\partial v/\partial x_g$. Thus, the higher x^* and the smaller the effect of x_g on the reputational value $v(\cdot)$, the more effective regulation of motives is in comparison to regulation of meaning. Of course, as the costs of both policies are not specified, this comparison cannot be made in absolute terms. The intuition for the result is that when almost everybody follows the norm, the marginal individual is rather insensitive to social pressure (i.e. a_i is small). Thus, changing the reputational value of the norm has little effect on the marginal individual's decision, and it is easier to influence this individual via monetary incentives.

Two examples of a reputational value function seem of special interest. I will analyze them in more detail below.[25]

3.1 Bandwagon norms

Bandwagon norms are characterized by the fact that once a critical proportion of the population follows the norm, it becomes very costly in terms of reputation not to follow it, which is illustrated in Figure 4. For a large number of norm followers, there is a diminishing marginal effect on reputation. Similarly, as long as only a few people exhibit a certain behavior, there is no effect on others in terms of reputation losses for not doing the same. Remember that the shape of the long–run equilibrium schedule is determined by those (x_t, μ_t)-pairs for which $v(x_t, x_g) = \mu_t$. Thus, the reputational value and therefore the equilibrium number of believers displays a steep increase in a critical range of norm followers, x_t. It may be that this property describes norm–following behavior better than the linear case in most instances. It approximates the extreme description that a norm is either perceived as valid or not by (almost) everybody in the group. Or it can describe a situation in which people want to do what the majority does. Consider the example of smoking norms which can exhibit critical mass effects. If the majority of people in a room or even in a society smoke, nonsmokers will typically not try to stop them. However, when the proportion of nonsmokers exceeds a certain critical point, they will feel more justified to ban smoking.

[25] I owe the general ideas of this section to the analysis of externalities in consumption, such as bandwagon and snob effects, by Leibenstein (1950) and Corneo and Jeanne (1997).

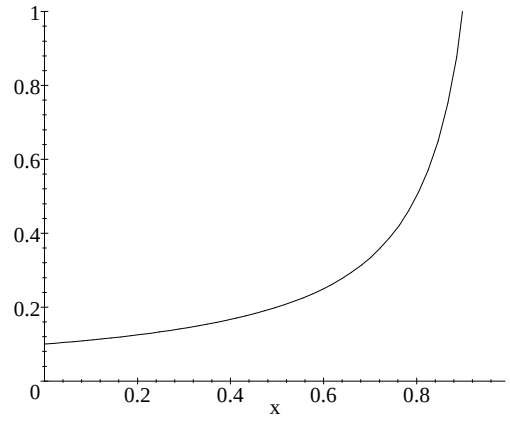

Figure 4: Bandwagon norms

What are the marginal effects of norm regulation on bandwagon norms? First consider an existing norm that is followed by almost everybody. Introducing incentives to break it has a strong effect on the number of norm followers as the schedule of the reputational value is almost flat in this upper portion. Thus, belief in the norm and its reputational value are almost not affected by m_g, which contrasts with the linear case. On the other hand, lowering the reputational value of the norm using x_g decreases x_t and μ_t along the short–run decision schedule, leading only to a small change in x_t. Thus, regulation of motives is at the margin more effective than regulation of meaning to change behavior.

The history of smoking regulation is rich and well studied.[26] The anti-smoking campaign beginning in the late 1960's relied on scientific findings about health effects of smoking, which were transformed into social meanings: the smoker as pariah, weak, reckless, without self-control. This was achieved by a mix of policies, among others by tying smoking to unhealthy behavior using advertisements and the ban of smoking in public spaces, i.e. a change in incentives. The model suggests that when a majority of people smoke, a change of incentives (like the prohibition to smoke in certain public areas or in restaurants) is most effective to discourage the marginal

[26] For an overview see Lessig (1995), p. 1025-34. In the U.S. cigarette smoking was first banned after women started smoking cigarettes at the end of the 19th century. However, this ban, which was meant to fix gender roles, was lifted after Word War I as men gave up pipes and cigars in favor of cigarettes during the war, for reasons of practicality.

individuals, i.e. those who are rather insensitive to social pressure and are most likely to stop smoking even if all others keep doing it. Returning to the example of gender-based labor supply norms, a change in the tax system is predicted to be more effective to increase labor market participation of women in Germany than symbolic changes, affecting the social meaning of the norm, like a few women in highly visible and important jobs.

Starting at the stable equilibrium with $x^* = 0$, regulation of motives to establish norm-following behavior by pushing the short–run decision schedule downwards must again be of sufficient size. But note that in contrast to the linear case, setting $v(0, x_g) \geq y + m_g$ may not be sufficient to get the bandwagon started. In particular, if $\frac{\partial \mu_t}{\partial x_t} > \frac{\partial v}{\partial x_t}$ for small x_t (i.e. the marginal increase in reputational value of one person following the new code is smaller than the private benefit from not following the norm), then it is necessary and sufficient for establishing a new norm that $v(x_t, x_g) \geq \frac{y+m_g}{1-x_t}$ for all x_t. In the case of new codes of ecological behavior, the monetary incentive to recycle empty bottles together with the increase in the reputational value of doing so (brought about by raising public awareness etc.) must be strong enough right away to set the process of norm formation in motion.

3.2 Snob norms

A snob norm confers the highest reputation to its followers when the number of those following the code is small. That is, the reputation gain from following the norm is negatively correlated with the number of followers once a certain number of followers x^s is reached. Thus, the reputational value of a snob norm peaks at this intermediate number of followers x^s as shown in Figure 5. An example for such a norm are group norms where certain actions qualify people as members of the in-group. If the group becomes too large, the social gains from adhering to it diminish. Different classes of society often develop such norms to differentiate themselves from others. For example huge money donations lead to reputation gains because (among other reasons) only a very limited number of people engages in this behavior. The fashion world abounds with examples for snob norms as there is a constant creation of new trends that increase the reputation of its followers (i.e. wearers) only as long as there are not too many of them. Now consider the effect of regulation of motives and meaning under this assumption.

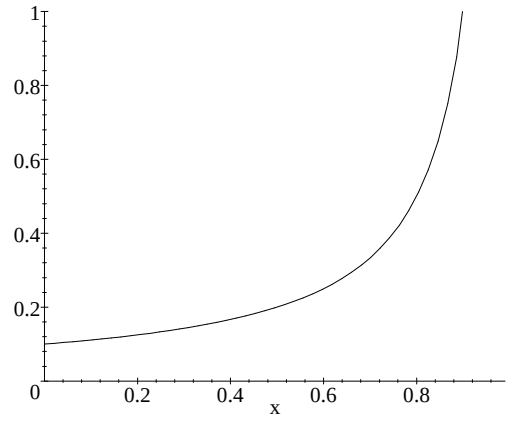

Figure 5: Snob norms

Suppose that a snob norm exists and that it is followed by a stable proportion of people, $x^* > x^s$. Introducing incentives for breaking it has the following effects: Although the number of followers is decreased by increasing m_g, the reputational value increases. This happens whenever the stable equilibrium is on the downward–sloping portion of the $v(x_t, x_g)$–schedule because a decrease in the number of followers benefits those adhering to the code by giving them an even higher reputation. This is the reverse of the linear case. A possible example is juvenile drug use. If using drugs confers social status or group membership, increasing the price of drugs or increasing the punishment decreases the number of drug users only by little, but increases the reputation gain of those still taking drugs.[27]

What are the effects of transforming the social meaning of the snob norm in order to destroy it? A snob norm can be discredited such that the reputation gain from following it is lower for every proportion of norm followers (e.g. if $v(\cdot)$ is linear in x_g). This amounts to shifting the curve downwards and leads to a reduction in the long–run equilibrium proportion of norm followers. But just as with regulation of motives, this decline of norm following behavior is small compared to bandwagon norms ceteris paribus. This can be taken from Proposition 4 as the denominator of the marginal effects is larger for snob norms than for bandwagon norms because $\partial v/\partial x_t < 0$. Thus,

[27] Another example is dueling, governed by a complicated set of norms, which is described as a practice that served to distinguish members of the upper class from others. Lessig (1995) argues that attempts by non-members of the elite to ban dueling increased solidarity and support for dueling in the American South.

destroying a snob norm is difficult, but should definitely not rely on monetary incentives alone. The model suggests for example that peer pressure to take drugs among young people can better be counteracted by information campaigns about the health hazards of taking drugs than by prohibiting and/or raising their price.

How can a snob norm be established that is non–existent, for example a norm requiring rich people to make huge money donations to public schools? Again, introducing incentives to follow this norm (tax exemptions for example), i.e. decreasing m_g to push down the short–run equilibrium schedule, can be successful if it is done to a sufficient degree. Alternatively, the norm can be established for example by Bill Gates announcing that he is giving a big sum of money to public schools. This can increase the reputational value of contributing by raising $v(x_t, x_g)$ sufficiently such that even if almost nobody does so (x_t close to 0), some people start following his example.

4 Conclusions

The process of norm formation and erosion is a product of both, collective action and individual attempts to challenge, support, discredit or introduce a certain code of behavior. This paper presents a simple model aimed at a better understanding of the interaction between these two elements of the dynamics of norms. In particular, it is an attempt to endogenize norms by investigating how individual actors or groups of individuals (norm entrepreneurs) can influence the set of relevant norms in a society. Two different ways of norm regulation are introduced, namely changing the incentive structure and changing the social meaning and reputational value of a norm.

The main result of the paper is that the optimal policy mix depends on the type of the norm under consideration. If a norm belongs to the class of bandwagon norms, erosion of the norm can be initiated most powerfully by changing the incentive structure. But destroying snob norms by changing the relative prices can be very costly compared to regulating their meaning by advertisement campaigns etc. Regarding the creation of new norms, the more important a critical mass as in bandwagon norms, the more people must be convinced simultaneously to adopt the new code. Creating snob norms on the other hand requires only a few individuals to adopt an action to make it attractive for others to take the same action because a small number of norm followers confers high reputational gains.

A number of questions remain unaddressed. Among them is a full welfare analysis, incorporating the costs and benefits of regulating norms.[28] A related question is how long a certain policy must be continued to establish or destroy a norm. Furthermore, it is conceivable that the full information assumption is often violated. Sometimes, norms can persist because people think that almost everybody else follows it, but when polls are published presenting the true figures of norm following behavior, the norm may unravel rapidly.

References

[1] Akerlof, G. (1980). A Theory of Social Custom, of Which Unemployment May Be One Consequence. Quarterly Journal of Economics XCIV: 749–75.

[2] Basu, K., Jones, E., and Schlicht, E. (1987). The Growth and Decay of Custom: The Role of the New Institutional Economics in Economic History. Explorations in Economic History 24: 1-21.

[3] Cooter, R.D. (1996). Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant. University of Pennsylvania Law Review. 144: 1643–1696.

[4] Cooter, R.D., and Ulen, T. (1997). Law and Economics (2nd ed.), Reading, MA: Addison–Wesley

[5] Corneo, G. (1995). Social Custom, Management Opposition, and Trade Union Membership. European Economic Review 39: 275–292.

[6] Corneo, G., and Jeanne, O. (1997). Snobs, Bandwagons, and the Origin of Social Customs in Consumer Behavior. Journal of Economic Behavior and Organization 32: 333–347.

[7] Elster, J. (1989). Social Norms and Economic Theory. Journal of Economic Perspectives 3:99–117.

[28] In the model presented above, maximizing a social welfare function, defined as the utility of norm breakers multiplied by the number of norm breakers minus the costs of regulation, is not a well behaved problem. The reason is that the variable m_g affects both the utility of individuals and the number of individuals breaking the code.

[8] Frey, B., and Kirchgässner, G. (1994). Demokratische Wirtschaftspolitik. München: Vahlen

[9] Hirschman, A. (1985). Against Parsimony. Economics and Philosophy 1: 7-21.

[10] Huck, S. (1998). Trust, treason, and trials: An example of how the evolution of preferences can be driven by legal institutions. Journal of Law, Economics,and Organization 14: 44-60.

[11] Kamiat, W. (1996). Labor and Lemons: Efficient Norms in the Internal Labor Market and the Possible Failures of Individual Contracting. University of Pennsylvania Law Review 144: 1953–1970.

[12] Leibenstein, H. (1950). Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. Quarterly Journal of Economics 64: 183–207.

[13] Lessig, L. (1995). The Regulation of Social Meaning. University of Chicago Law Review 62: 943–1045.

[14] Naylor, R. (1989). Strikes, Free Riders, and Social Customs. Quarterly Journal of Economics CIII: 771–85.

[15] de Neubourg, C., and Vendrik, M. (1994). An Extended Rationality Model of Social Norms in Labour Supply. Journal of Economic Psychology 15: 93–126.

[16] Posner, E. (1996). Law, Economics and Inefficient Norms. University of Pennsylvania Law Review 144: 1697–1744.

[17] Romer, D. (1984). The Theory of Social Custom: A Modification and Some Extensions. Quarterly Journal of Economics 99: 717–27.

[18] Rock, E.B., and Wachter, M.L. (1996). The Enforceability of Norms and the Employment Relationship. University of Pennsylvania Law Review 144: 1913-1952

[19] Schlicht, E. (1998). On Custom in the Economy. Oxford: Clarendon Press

[20] Schäfer, H.-B., and Ott, C. (1993). Emergence and Construction of Efficient Rules in the Legal System of German Civil Law. International Review of Law and Economics 13: 285–302.

[21] Sugden, R. (1989). Spontaneous Order. Journal of Economic Perspectives 3:85–97.

[22] Sunstein, C. (1996). On the Expressive Function of Law. University of Pennsylvania Law Review 144: 2021–2053.

[23] Young, H.P. (1998). Social Norms and Economic Welfare. European Economic Review 42: 821–830.